



**Technical contribution within the framework of
the regional thematic hearing of the
Inter-American Commission on Human Rights
Humans**

**“Artificial Intelligence and Human Rights
Humans”**

Submitted by the Electronic Frontier Foundation and
ARTICLE 19 Office for Mexico and
Central America

About Article 19 Office for Mexico and Central America

ARTICLE 19 Office for Mexico and Central America (ARTICLE 19) is an independent non-governmental organization that promotes the right of all people to information and expression in an environment of freedom, security, and equality, both online and offline .

The organization began working on Mexico's artificial intelligence agenda more than four years ago. The goal was to understand how the initial efforts in its development, implementation, and regulation originated, the obstacles and challenges that arose, and the degree of political will to adopt this agenda as a political priority—given the evolution of technologies and their incorporation into various spheres of public life.

The findings and conclusions of our initial exploration revealed a lack of interest in discussing emerging technologies within the public sector with the seriousness it demands. However, the new Mexican government administration has recently demonstrated its interest in addressing the technological issue, beginning with reducing the digital divide through infrastructure and simultaneously promoting the digitalization of the public sector to carry out various processes, such as administrative procedures and the administration of justice.

However, at ARTICLE 19, we have currently observed a lack of understanding of what artificial intelligence is and its potential impact on human rights. This occurs when legislative initiatives inhibit the interdependence of human rights. Furthermore, ARTICLE 19 has analyzed the lack of depth and understanding of algorithmic transparency, highlighting the need to address its structure, challenges, and the priority of linking it to principles and guidelines related to access to information, transparency, and accountability. This is directed not only at the private companies developing these tools, but also at the governments that are using these technologies within their national digitalization strategies.

In this regard, ARTICLE 19 has been invited as an expert organization to participate and facilitate discussions with other stakeholders, contributing to these conversations with a human rights perspective and focusing on issues of freedom of expression and access to information. ARTICLE 19 works to link public policy advocacy, local process support, and strategic litigation to promote the highest international human rights standards in the development, use, acquisition, and

implementation of emerging technology tools, in such a way that it contributes to strengthening democracy.



About the Electronic Frontier Foundation

The Electronic Frontier Foundation (EFF) is a U.S.-based, nonprofit, non-governmental organization operating internationally to protect human rights and fundamental freedoms in our digital age. Founded in 1990, EFF defends privacy, free expression, and innovation through strategic litigation, public policy analysis, grassroots activism, and technology development. EFF's mission is to ensure that technology supports freedoms, justice, and innovation for all people.

As part of its work, the EFF has been monitoring initiatives to regulate artificial intelligence in Latin America and the deployment of these technologies for sensitive uses, particularly by state institutions for activities and decisions that affect human rights. The lack of meaningful transparency and citizen participation in the way these systems are generally developed, implemented, and used by state institutions are key factors that pose serious challenges to democratic governance and the legitimacy of decision-making and government policymaking. Based on Inter-American human rights standards, we have developed implications that states should consider from the moment they analyze whether it is appropriate to adopt an algorithmic system as part of a given state action to the moment they evaluate the results of its implementation.

In connection with the regional thematic hearing entitled *Artificial Intelligence and Human Rights*, which took place on Friday, March 7, 2025, ARTICLE 19 and EFF submit these technical contributions with the purpose of analyzing the challenges that artificial intelligence (AI) poses for the exercise and guarantee of human rights, and offering some recommendations to prevent and mitigate them.

A. Artificial intelligence and its relationship with the exercise and guarantee of protection of rights human rights

1. Automated content moderation on social media platforms

AI has positioned itself as a catalyst for transformation¹ across all disciplines and industries, reconfiguring traditional processes. This is why AI has been integrated into diverse disciplines and practices, not only in law but also in areas such as health, education, geography, and many others. For example, in academia, AI can enhance research and teaching, facilitating the analysis of large volumes of information and optimizing the preparation of legal texts.

The Internet is no longer just a means of disseminating and publishing information; it is also a necessary tool for the exercise of human rights. Thus, technology becomes an even more necessary tool as part of work, family, and social life, education, expression, and the construction of our identity in the social sphere.

collective as well as individual.

An example of this is how social media has become a space for discussion, knowledge sharing, reflection, disputes, and even narrative contests in forums where many people discuss various topics.

Like any other social space, the Internet has also changed the dynamics of virtual interaction, reproducing stereotypes, roles, and social practices from offline spaces. This is evidenced by the violence, inequality, and injustice reflected in digital spaces, as well as the growing economic, educational, and health gaps caused by inequality based on gender, racial, ethnic, and social class in our communities.

Due to certain practices on social media and digital platforms, excessive violence has been witnessed, as well as exacerbating inequality. Digital platforms have created mechanisms to control or moderate content by classifying, at their own discretion, speech, audio-visual content, and expressions as harmful and/or problematic when they violate the rules, regulations, or terms of service of their platforms.

¹ <https://www.article19.org/wp-content/uploads/2018/04/Privacy-and-Freedom-of-Expression-In-the-Age-of-Artificial-Intelligence-1.pdf>

2. AI as a tool to enhance democracy? Risks and opportunities of AI

Information technology (IT) in the government context is critical to driving operational efficiency and facilitating data-driven decision-making. It encompasses a range of solutions, including computing systems, software, and networks, used to effectively process and distribute information. These technologies enable governments to improve the delivery of public services by automating tasks, processing large volumes of information, and communicating data.

B. Identification and prevention of differentiated impacts on vulnerable groups

~~The use of AI in vulnerable and marginalized groups~~

The report [Artificial Freedom: Content Moderation and Its Impact on Human Rights](#) Prepared by ARTICLE 19, , addresses how automated content moderation in Digital platforms can affect the right to freedom of expression and other fundamental rights. This is based on a documentary analysis, surveys, interviews, and workshops with vulnerable communities in Mexico.

The study highlights that content moderation by digital platforms, particularly when done automatically, can lead to censorship, discrimination and self-censorship, differentially impacting different sectors of society that fear [retaliation](#). of governmental and/or non-state actors³ . In relation to

Given the above, digital platforms do not always effectively control so-called hate speech, which particularly affects women, journalists, human rights defenders, LGBTQ+ people, and migrants, who are more likely to suffer harassment and digital violence. As a result, people with disabilities are often stigmatized, and their content is often misinterpreted or eliminated by algorithms.

In addition, the report shows evidence of how automated content moderation on digital platforms can reinforce structural inequalities and restrict fundamental rights such as freedom of expression and access to information.

² ARTICLE 19 Mexico and Central America Office. *Artificial Freedom: Content Moderation and Its Impact on Human Rights*. October 19, 2022. <https://articulo19.org/libertad-artificial-discurso-redes-y-pluralidad/>

³ ARTICLE 19 Mexico and Central America Office. *Pending Rights*. July 9, 2024. <https://articulo19.org/derechospendientes/>

In conclusion, the document urgently calls for transparency, appeal mechanisms, and human rights-based regulation to ensure that digital platforms do not perpetuate discrimination and censorship in the digital space.

Some of the impacts that stand out are:

1) Freedom of expression and access to information

- Removing legitimate content limits the diversity of voices in the digital space.
- Public debate is restricted on key issues such as human rights, politics and social activism.
- A self-censorship effect is generated: people stop expressing certain opinions for fear of being blocked.

2) Rights of vulnerable groups

- Indigenous and Afro-Mexican communities: There are language and cultural barriers that prevent algorithms from correctly interpreting their content.
- LGBTQ+ people: They have faced unjustified removals of content related to their identity and rights.
- Women: The invisibility of reports of gender violence has been detected, while digital attacks against them persist without an adequate response.

3) Digital privacy and security

- Platforms collect large volumes of data without ensuring transparency in its use.
- There are no clear guarantees about how the generated data is stored or used by users in content moderation processes

Use of emerging technologies and lack of algorithmic transparency in the public sector

ARTICLE 19 published the report entitled "[Towards algorithmic transparency in the Latin American public sector.](#)" This study addresses the challenges of transparency in the acquisition, development, and implementation of emerging technologies in public administration in Argentina, Brazil, Colombia, Chile, and Mexico. The study identifies the lack of transparency in the use of artificial intelligence (AI) algorithms and systems in the public sector, highlighting three key problems:

- A. **Lack of regulation on algorithmic transparency:** There are no specific regulations in the region that require governments to explain how AI systems used in administrative decision-making work;
- B. **Difficulty in supervision and accountability:** The public administration lacks clear mechanisms to monitor the impact of AI and ensure that its decisions are auditable;
- C. **Impact on human rights and citizen participation:** The lack of transparency in the implementation of emerging technologies prevents citizens from questioning or appealing automated decisions.

One of the main findings relates to the lack of access to information on AI in the public sector. It is important to note that none of the five countries studied has a centralized registry of AI systems used in public administration.

Furthermore, contracts and tenders related to emerging technologies are often protected by confidentiality agreements, which prevents their disclosure. For these types of situations, no clear standards were found to ensure that the algorithms used are explainable and accessible to the public.

Purchase, acquisition and use of emerging technologies by the public sector (supply chain)

The acquisition and deployment of AI systems by states presents significant risks to human rights in a variety of areas, with profound implications for privacy, equality, democratic accountability, and public participation in digital innovation systems.

(i) Law Enforcement: The use of AI in law enforcement, such as predictive policing and facial recognition technology, raises significant human rights concerns, including biased outcomes that disproportionately harm marginalized communities, lead to unjust incarceration, and erode people's privacy through extensive data collection.⁴ Although technologies such as facial recognition may have a low error rate in absolute terms, their deployment still results in large numbers of people being wrongly tracked, amplifying the surveillance of people's movements and lives. These practices threaten individual autonomy, stifle freedom of expression, and threaten to create a state of

⁴ <https://link.springer.com/article/10.1007/s13347-021-00478-z>

surveillance.⁵ The opacity of these AI systems makes accountability even more difficult law enforcement.

(ii) Social assistance: AI-based social assistance systems risk unfair exclusions, surveillance, and lack of accountability.⁶ The arbitrary and unfair denial of social benefits⁷ deprives people of their right to life, equality, an adequate standard of living, and subsistence.⁸ When public services, such as resource allocation, are not transparent, this limits opportunities for public participation in policy development and evaluation, and hampers individuals' ability to seek redress for bias or other harm. This is particularly true when private actors are involved in performing public service functions.

(iii) Education: AI in education poses risks such as omnipresent surveillance, behavior-based rewards or punishments, and the stifling of critical thinking.⁹ It can exacerbate violence, harassment, and cyberbullying.¹⁰ A UN Special Rapporteur highlights the risks of using AI for admissions, student allocation, performance assessments, and behavior monitoring during exams.¹¹ AI could also amplify educational inequalities, widening the gaps between privileged and disadvantaged students, as AI tools often favor dominant cultures and languages.¹² Also of concern are reduced human interaction, a declining role for teachers, technocratic approaches, and the exploitation of student data for non-consensual or socially hostile purposes.¹³

(iv) Child protection services: States are increasingly automating processes for predicting the risk of violations of children and adolescents' rights. The growing use of AI to predict risks to children's rights allows private sector actors to access and exploit highly sensitive data without adequate privacy protections or safeguards against abuse and exploitation.¹⁴

⁵ <https://articulo19.org/continua-impune-el-uso-de-pegasus-a-uno-de-las-nuevas-denuncias/>

⁶ <https://www.amnesty.org/en/latest/news/2024/11/denmark-ai-powered-welfare-system-fuels-mass-surveillance-and-risks-discriminating-against-marginalized-groups-report/>

⁷ <https://www.amnesty.org/en/latest/research/2024/04/entity-resolution-in-indias-welfare-digitalization/>

⁸ <https://ohrh.law.ox.ac.uk/the-human-rights-implications-of-chinas-social-credit-system/>

⁹ <https://www.mdpi.com/2071-1050/15/16/12451>

¹⁰ <https://slejournal.springeropen.com/articles/10.1186/s40561-024-00316-7>

¹¹ <https://documents.un.org/doc/undoc/gen/n24/298/43/pdf/n2429843.pdf>

¹² https://www.researchgate.net/publication/351119867_Equity_and_Artificial_Intelligence_in_Education_Will_AI_Ed_Amplify_or_Alleviate_Inequities_in_Education

¹³ <https://documents.un.org/doc/undoc/gen/n24/298/43/pdf/n2429843.pdf>

¹⁴ <https://www.mdpi.com/2071-1050/15/16/12451>

(v) Judiciary: Since AI tends to reinforce structural biases, its use in the judiciary, especially for risk assessment in sentencing, can produce discriminatory outcomes,¹⁵ undermining the right to a fair trial and treatment, as well as compromising judicial discretion.¹⁶ A recent ruling by a judge in Colombia to protect the right to health of a minor with autism, in which the legal reasoning was entirely drafted by ChatGPT, raises concerns regarding judicial independence and the duty to issue reasoned judgments.¹⁷

(vi) Migration and Border Management: AI technologies in migration systems increasingly threaten migrants' human rights and their access to asylum.¹⁸ A 2024 Electronic Frontier Foundation report highlights how layered surveillance technologies, frequently automated by AI, control border communities.¹⁹ The Sentinel Platform in Chihuahua demonstrates how opaque systems, including biometric identification and behavioral prediction, are used under the guise of public safety and immigration enforcement, raising concerns about transparency, accountability, and human rights impacts.

humans.

(vii) War and conflict: AI in military applications poses serious risks to civilian life and accountability. Autonomous weapons and AI-based surveillance systems reduce human oversight and blur accountability.²⁰ In 2024, UN experts condemned the use of AI²¹ and called for restorative measures, highlighting the ethical and humanitarian implications of such technologies.²²

The above examples demonstrate how critical state functions, such as welfare, education, justice, and the maintenance of the public commons, are threatened with erosion by the slow advance of unaccountable private interests in the public sphere.²³ The leakage of data value from the public sphere to ecosystems

¹⁵ <https://researchoutreach.org/articles/justice-served-discrimination-in-algorithmic-risk-assessment/>

¹⁶ https://www.unodc.org/dohadeclaration/en/news/2021/30/judge-dread_ai-and-judicial-integrity.html

¹⁷ <https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling>

¹⁸ <https://www.amnesty.org/en/latest/news/2024/05/global-new-technology-and-ai-used-at-borders-increases-inequalities-and-undermines-human-rights-of-migrants/>

¹⁹ <https://www.eff.org/files/2024/05/06/borderzine-2024-5-6-es.pdf>

²⁰ <https://hms.harvard.edu/news/risks-artificial-intelligence-weapons-design>

²¹ Domicide refers to the systematic or widespread violation of the right to adequate housing; See <https://documents.un.org/doc/undoc/gen/n22/431/55/pdf/n2243155.pdf>

²² <https://ainowinstitute.org/publication/the-algorithmically-accelerated-killing-machine>

²³ <https://itforchange.net/sites/default/files/2647/Recovering%20the%20Public%20in%20India%E2%80%99s%20Digital%20Public%20Infrastructure%20Strategy.pdf>

captives of the private sector points to the need to democratize data and AI dividends so that they serve the collective good.

In Mexico, however, the disappearance of the Institute for Access to Information and Transparency (INAI) reveals a clear institutional weakness in the face of an imminent risk to accessing information, where 90% of technology acquisitions are made through direct procurement, limiting competition and increasing the risk of corruption, even opening up the possibility of human rights violations.

Nayarit Prosecutor's Office Case Study, Mexico. Diagnostic of the Use and Implementation of AI Technologies in the Administration of Criminal Justice.

The document titled "[Reflections on the Implementation of AI Systems in the Administration of Justice. Case Study: Nayarit](#)" It analyzes the challenges, opportunities, and risks of integrating artificial intelligence into the criminal justice system, particularly in Mexico, taking the state of Nayarit as an example.²⁴ This report proposes a framework for implementing AI in the administration of justice in an ethical, transparent, and human rights-respecting manner, where critical areas for improvement are identified.

The document mentions everything from poor attention to technological infrastructure (up to 26.32%) to lack of staff training (includes 87.72%) which can lead to risks such as biases in algorithms, transparency problems and the protection of personal data where 98.25% of staff are unaware of ARCO rights (access, rectification, cancellation and opposition) and inadequate file management and cybersecurity where 77.19% of staff are unaware of cybersecurity systems, in addition to the existence of insecure practices such as sharing passwords and using external devices without control.

The findings include that most people are not trained in AI, that equipment is outdated, and that data protection regulations are not complied with. It also notes that 82.45% of people do not use online collaboration tools, indicating a significant digital divide of 94.74%. In addition, the lack of key documents such as privacy notices and risk analyses demonstrates a lack of compliance with laws, which could lead to sanctions or even perpetuate human rights violations.

²⁴ <https://articulo19.org/nayarit-case: Reflections on the Implementation of AI Systems in the Prosecution Service>

The impact on human rights assessed with the information collected was as follows:

1. Privacy and data protection:

- o The absence of robust measures exposes personal data of victims, witnesses and accused to leaks or misuse, violating the right to privacy (Art. 16 of the Constitution).

2. Non-discrimination and equality:

- o Biases in algorithms could replicate discriminatory patterns (e.g. racial or socioeconomic profiles), affecting vulnerable groups and violating the principle of equality (Article 1 of the Constitution).

3. Due process:

- o Automated decisions without human supervision could lead to errors in charges or precautionary measures, compromising the right to a fair trial (Article 20 of the Constitution).

4. Access to justice:

- o The lack of transparency in the use of AI erodes citizen trust, making effective access to justice difficult.

C. The implementation of effective protection mechanisms

State mechanisms addressing the negative impacts of AI on human rights generally fall into three categories:

(i) human rights (HRI) on

HRIAs by states and companies can serve as an empowerment tool, providing affected individuals with information to challenge the design, deployment, and operation of AI systems. For example, Article 72 of the EU AI Law mandates fundamental rights impact assessments for high-risk AI systems, enabling transparency and accountability.

(ii) Notification and fair hearing before decisions adverse:

People must be informed when an automated system is used in decisions that affect them. The US Blueprint for the AI Bill of Rights emphasizes timely and understandable notification, opt-out options, and explanations about the process.

Key functional changes.²⁵ Individuals must also have access to human oversight and appeal mechanisms before adverse decisions are made. These mechanisms must be accessible, equitable, effective, and timely, and must not impose an undue burden on the affected individual. Other examples include New Zealand's Algorithms Charter, which mandates appeals of automated decisions,²⁶ and 24 states' ballot curation laws in the United States, which provide alternative systems for voters flagged by signature-matching algorithms.²⁷

(iii) Public algorithm repositories: Within the public sector, the creation of public algorithm repositories can improve explainability and accountability by opening up the black box of AI development and deployment. Even where full algorithm transparency may not be desirable, repositories can contain use cases where the algorithms are used, along with explanations of their role in the decision-making process, ensuring transparency and enabling remediation of adverse impacts.²⁸ Such repositories can also be part of a broader algorithmic transparency strategy, extending to the private sector and integrating HRIA, grievance redress, and reporting requirements.

(iv) Judicial remedies: Courts can play a critical role in addressing AI-related injustices. To improve access to justice, equality bodies and public interest organizations should be empowered to file complaints, even without identifiable complainants, as AI-driven rights violations often go unnoticed by victims.

Reviewing the rules of evidence is also crucial to ease the burden of proof on plaintiffs, given the complexity and opacity of AI systems. The EU Directive on AI Liability proposes a presumption of causality between breach and damage.²⁹

Examples of court cases:

I The Dutch SyRI case (2020): The District Court of The Hague ruled that the System Risk Indication (SyRI) algorithm system, a legal instrument that the

²⁵ <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

²⁶ http://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

²⁷ <https://www.lawfareblog.com/mail-voting-litigation-2020-part-iv-verifying-mail-ballots>

²⁸ <https://wp.oecd.ai/app/uploads/2024/12/16-Algorithmic-Transparency-in-the-Public-Sector-Recommendations-for-Governments-to-Enhance-the-Transparency-of-Public-Algorithms.pdf>

²⁹ [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI\(2023\)739342_ES](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739342/EPRS_BRI(2023)739342_ES).

The Dutch government uses this information to detect fraud in areas such as benefits, subsidies, and taxes, which violates the right to privacy under the European Convention on Human Rights.³⁰

Mobley v. Workday (2023): A US court granted leave to hear a lawsuit against Workday's AI-powered hiring tools, which allegedly discriminated on the basis of race, age, and disability. The ruling highlighted that AI providers can be held liable under anti-discrimination laws when acting as agents of employers.³¹

D. Compliance with state obligations based on inter-American standards.

Looking at the commitments of the States of the region to human rights, the Electronic Frontier Foundation (EFF) recently published [a report](#) on the state use of systems based on algorithms and artificial intelligence (AI) to automate tasks and procedures in support of functions and decisions that affect rights. The report highlights that, although the deployment of these technologies by public entities may be novel, the foundation of rights guarantees on which their use must be based dates back many years and includes robust standards that States must guarantee.

It should be remembered that the American Convention on Human Rights establishes [the obligation](#) of the States Parties to respect and guarantee [human rights](#), which include the duties to prevent, investigate and punish violations of said rights. In this context, and taking into account previous sections of this contribution, it is key to emphasize that the adoption of algorithmic systems as part of policies or initiatives of State bodies entails security risks, discrimination, illegal surveillance and errors that can affect access to essential services or lead to serious consequences for the lives of thousands of people.

people.

The implementation of technology in the public sector therefore requires a careful examination by the bodies that contemplate the acquisition or use of these technologies to determine whether the use of an AI system is appropriate for the intended purpose, as well as whether it offers a concrete benefit. These purposes and benefits must be associated with the strengthening of the guarantee and exercise of certain rights, which must be verified when monitoring and evaluating its implementation. Thus, the adoption

³⁰<https://www.loc.gov/item/global-legal-monitor/2020-03-13/netherlands-court-prohibits-governments-use-of-ai-software-to-detect-welfare-fraud/>

³¹<https://www.seyfarth.com/news-insights/mobley-v-workday-court-holds-ai-service-providers-could-be-directly-liable-for-employment-discrimination-under-agent-theory.html>

of these systems by state entities requires the performance of human rights impact assessments before deployment and during use.

In addition, any human rights violations arising from the adoption of these systems and technologies by public institutions must result in investigations that provide input for the evaluation of the system and the context of its use, leading to effective modifications or the discontinuation of its use. This also develops the necessary measures to prevent the rights violations from recurring.

Respecting and guaranteeing rights implies organizing the entire apparatus of public power to give effect to the rights established by the Convention, including adopting legislative measures and adjusting State practices, processes, and structures toward this objective. All State institutions are subject to the commitments assumed under international human rights law. This includes the judges and courts responsible for exercising conventionality control over national laws and regulations that

base state use of AI systems or automated decisions on situations or through unlawful procedures under the American Convention.

Cross-cutting principles

A fundamental element emerging from inter-American guarantees is the imperative of transparency in state activities. The principle of maximum disclosure applies to them, according to which the right of access to information is the rule and secrecy is the exception. The State bears the burden of proving that access to certain information can be legitimately restricted. Furthermore, the right to information takes precedence over conflicting norms or the lack of regulation.

Broad restrictions based on a general classification of “public order,” “national security,” or “public safety” are not permitted. Limitations must be strictly applied to the types of information and disclosures that, under the law, actually pose a risk based on democratic principles, and only if such a specific and objective risk exists. The requirement for transparency is reinforced in initiatives involving large amounts of public funding, which is often the rule in the deployment of AI technologies by state entities.

Due process guarantees in the implementation of AI technologies must be part of the essential foundations for their adoption for decision-making that affects rights. People affected by automated decisions must be informed of the procedures that analyze their rights and obligations before a decision is made. They must receive, with sufficient notice to allow them to intervene,

Clear information explaining why they are subject to the procedure, the relevant elements being considered, and a reference to how the elements being assessed influence the consequences they may face. The necessary information also includes the means available to them to present their arguments, which must be easily accessible. Furthermore, decisions must be duly justified, which requires the systems used to observe standards of interpretability and explainability.

Added to these guarantees is the right to privacy in its relation to data protection and self-determination. In a [recent decision](#) The Inter-American Court of Human Rights has recognized the protection of informational self-determination under the American Convention. This guarantees a series of rights linked to the guaranteed right of individuals to know and influence how their personal data is processed by third parties.

It is also essential to remember that the American Convention prohibits all discriminatory treatment. Consequently, before implementing algorithmic systems and throughout their lifecycle, States must take all necessary measures to ensure that they do not result in discriminatory treatment—that is, any exclusion, restriction, or privilege that is not objective, necessary, and proportionate to the legitimate aim they seek to achieve and that negatively affects human rights. States are therefore required to adopt practices to fulfill this obligation. These practices include the aforementioned human rights impact assessments and other measures such as: ensuring that responsible State institutions and oversight bodies have access to the internal workings of the system and the appropriate expertise to assess it from a human rights perspective; conducting independent internal and external audits; and ensuring meaningful human oversight and review of AI-based decisions by competent officials following strict and transparent criteria.

Periodic evaluation, impact assessments, and audits should address not only the initial design, but also the actual results of the use of algorithmic systems once implemented. Human rights impact assessments and audits should adopt a gender and intersectional perspective in accordance with the three steps that the IACHR has identified for the design of public policies with a human rights approach.³² These steps constitute a relevant roadmap for States to evaluate public policies that make use of algorithmic systems. They include consideration of: (i) the differential impact that a given policy has or may have on historically discriminated groups; (ii) the perspectives and concerns of these groups;

³² IACHR, Public Policies with a Human Rights Approach, September 15, 2018, para. 48.

groups at different stages of the policy cycle; (iii) and the real benefits that this policy brings or can bring to reducing the inequalities affecting these groups.

It should be emphasized that it is the responsibility of State authorities, and not the affected individuals, to justify and demonstrate that a decision based on automated systems or AI models did not have a discriminatory purpose or effect. To properly comply with this obligation, States must ensure that decisions based on these systems have explainable and justifiable reasons. Participation mechanisms can take multiple and complementary forms. The [operational framework](#) The report prepared by the EFF articulates some of [these mechanisms](#) and structures, including system audits in collaboration with independent data scientists and civil society experts, broader consultation processes, feedback and complaint mechanisms, citizen participation within oversight institutions, as well as coordination with public bodies that advocate for the interest of society (such as ombudsmen's offices, consumer protection bodies, among others).

Finally, meaningful social participation is another key element of state action based on human rights. This presupposes access to relevant information on the use of these systems and the involvement of various sectors of society in the decision-making and evaluation of their implementation, especially historically discriminated groups, affected individuals, and experts on the topics related to each of these initiatives (e.g., education, social protection, etc.).

Transparency and access to information, due process, equality and non-discrimination, privacy and data protection, and meaningful social participation are all cross-cutting principles that should guide the State's actions in the deployment of AI systems in the performance of its functions—not only to guide **how** to do so and evaluate its implementation, but also to base the initial definition of **whether** such deployment is legitimate and in accordance with human rights.

Red lines and tripartite test for the restriction and limitation of human rights

It is crucial to emphasize that the deployment of algorithmic systems to support the performance of public functions is not always an appropriate approach to addressing the problems they seek to address. While States' use of algorithmic systems may, in theory, serve to advance the exercise of human rights, supporting sensitive decision-making within the State function in these systems carries, at the very least, the possibility of interfering with one or more of those rights and freedoms. Consequently, the development or adoption of these systems by States must respect the Convention's grounds for permissible restrictions or limitations on human rights, which is the tripartite test. This means that, in addition to an adequate legal basis, the restrictions or limitations must pursue a legitimate aim and be appropriate, necessary, and proportionate.

to achieve this goal, ensuring that the State's duty of protection is fully fulfilled.

In the case of social, economic, and cultural rights (ESCR), such an analysis must integrate the principles of progressive development and non-regression. The adoption by States of algorithmic systems in the context of social protection, or welfare, must promote the adequate enjoyment and attainment of ESCR rights and must not entail a regression in social protection guarantees and policies. Any restriction on ESCR rights must also be justified and balanced in light of these two principles.

In this regard, there are important red lines that state institutions must observe when evaluating and deciding whether the adoption of an algorithmic system is truly appropriate for the intended purposes. Among other things, the use of these systems is *not* appropriate *when*:

- the decision must be based on human prudence, reasoning or experience and automatic pattern recognition in available data sets has no useful role to play in informing human intervention;
- the goal pursued is difficult to translate into mathematical variables (e.g., population happiness), and there are no suitable indicators that can be measured from the available data. Even if some *proxies* could work, the mathematical translation of a complex social phenomenon and the use of related *proxies* may not be the most appropriate approach to addressing the issue or problem;
- the current state of the technology (and of the specific system being considered for adoption) falls short of addressing the safeguards and rights that should be applied taking into account the purpose of using the system (e.g., the adoption of a non-explainable system to support risk assessment by an administrative or judicial authority);
- System developers and providers do not offer sufficient guarantees that these are reliable and secure, aligned with human rights standards that include transparency, non-discrimination, privacy, security, among other attributes;
- the system must present a history of human rights violations;
- developers and providers prevent public institutions in charge of system deployment and independent auditors from accessing the system's source code and executables, anonymized training data sets and testing materials, including anonymized test data sets;
- The sole purpose of adopting the system is to reduce costs. The legitimate purpose pursued by adopting the system must be related to the protection and/or promotion of human rights beyond mere cost reduction.

States should refrain from making decisions based on algorithmic systems in contexts that would be incompatible with human rights. That is, purposes, contexts, or systems that:

- are based on State practices that reproduce systemic discrimination and/or lead to racial profiling (e.g., the use of facial recognition systems for public safety and crime prediction systems);
- have a disproportionate impact on specific groups and/or inherently reproduce discriminatory views or practices reflected in biased data sets used to train the AI model or feed the system's operation, or that otherwise result from or reproduce prejudices or stereotypes (the examples noted in the previous point, as well as social protection systems/policies that single out specific groups as potential fraudsters);
- are configured as indiscriminate surveillance of physical or virtual spaces;
- imply a regression in social protection guarantees;
- represent the “instrumentalization” of individuals (e.g., exploiting the data of the most vulnerable or testing experimental solutions on marginalized populations as incentives for collaboration with the private sector).

D. Regulatory and institutional frameworks and the promotion of new developments normative.

Effective public procurement frameworks are vital to ensuring the responsible use of AI, fostering trust, shaping the ethical development of AI, and ensuring increased net public value.

Examples on an international scale:

At the international level, UN Resolution A/HRC/50/56 provides guidelines for the acquisition, purchase and licensing of AI for the provision of public services.³³

UNESCO's 2021 Recommendations on the Ethics of Artificial Intelligence represent the first global agreement outlining shared values and principles to guide the development of legal frameworks for the ethical and sustainable advancement of AI that align with human rights and the SDGs.³⁴

The World Economic Forum's *AI Procurement in a Box* paper outlines key considerations for governments before procuring and deploying AI solutions, including the importance of early impact assessments.

³³ <https://documents.un.org/doc/undoc/gen/g22/323/96/pdf/g2232396.pdf>

³⁴ <https://www.unesco.org/en/artificial-intelligence/recommendation-ethics>

AI, data sensitivity, data quality, and consent. It also considers the socioeconomic impact, especially on vulnerable populations.³⁵

The report of the United Nations General Assembly of the Special Rapporteur (A/79/520) on the right to education highlights how AI challenges the fundamental pillars of education - curriculum, pedagogy and assessment - and proposes a human rights-based approach that emphasizes participation, accountability, non-discrimination and transparency to guide the integration of AI in education.³⁶

It is also important to refer to existing sectoral principles and frameworks, such as environmental and health standards, to uphold established guiding principles on various aspects now affected by AI. For example, the UNECE Aarhus Convention³⁷ grants citizens rights to information and participation in environmental matters, while the Nagoya Protocol guarantees equitable access to and benefit-sharing arising from the utilization of genetic resources, which is relevant given the increasing integration of AI into agriculture, food systems, and genetic sequencing.³⁸

Example at the national level

The EU AI Law is a landmark piece of legislation that takes a risk-based approach, prohibiting high-risk AI applications, such as manipulative or exploitative systems, while imposing strict governance, risk management, and transparency requirements. ³⁹ The US AI Bill of Rights Project provides an important framework for government, technology companies, and citizens to work together to ensure more responsible AI. ⁴⁰

In October 2024, the White House released a memo on advancing responsible AI procurement in government, outlining new contracting requirements for federal agencies.⁴¹

³⁵https://www3.weforum.org/docs/WEF_AI_Procurement_in_a_Box_AI_Government_Procurement_Guidelines_2020.pdf

³⁶<https://www.ohchr.org/en/documents/thematic-reports/a79520-artificial-intelligence-education-report-special-rapporteur-right>

³⁷ <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction>

³⁸ <https://www.cbd.int/abs/default.shtml>

³⁹ <https://artificialintelligenceact.eu/the-act/>

⁴⁰ <https://www.whitehouse.gov/ostp/ai-bill-of-rights/>

⁴¹<https://www.cov.com/en/news-and-insights/insights/2024/10/omb-releases-requirements-for-responsible-ai-procurement-by-federal-agencies>

In New Zealand, the Minister of Statistics launched the Algorithms Charter, committing government agencies to improve transparency and accountability in their use of algorithms.⁴² Brazil's AI law grants individuals and groups affected by high-risk AI systems fundamental rights, such as the right to an explanation of the system's decisions, the right to challenge and request reviews, and the right to human oversight of decisions.

E. General recommendations:

a) On automated content moderation. To mitigate the negative impacts of automated moderation, the report proposes several strategies:

1. Greater transparency and accountability

- Platforms must publish the criteria used in a clear and accessible manner for content moderation.
- Platforms are encouraged to conduct external audits of their systems. automated moderation to assess their impact on human rights.

2. Implementation of effective appeal mechanisms

- Users should have the right to challenge content removal decisions.
- Platforms must have human reviews in cases where the freedom of expression.

3. Regulation based on human rights

- A regulatory framework is needed to ensure that content moderation respect freedom of expression and other fundamental rights.
- Governments should develop standards in line with the Inter-American framework of human rights.

⁴² https://data.govt.nz/assets/data-ethics/algorithm/Algorithm-Charter-2020_Final-English-1.pdf

4. Including diverse perspectives in algorithm design and improving moderation automated.

- The participation of vulnerable communities in the creation of content moderation policies is essential.
- Algorithms must consider the cultural and linguistic context to avoid unfair content removal. This highlights the need for the human review mechanisms already created by digital platforms to be strengthened, transparent, and, above all, offer the possibility of effective, simple, and rapid appeals.

b) On human rights related to the acquisition and deployment of AI by public power

The importance of an underlying structure for the protection of rights in accordance with the cross-cutting principles presented above

We believe there is a set of rights that represent principles that States must observe, from the initial assessment of the adoption of algorithmic systems for purposes that affect human rights, through the monitoring and evaluation phase of these policies and projects. These cross-cutting principles correspond to a fundamental framework that States must have in place, or have effective and advanced measures toward establishing them, for the proper deployment of such systems. Having this framework is not secondary; it derives from States' obligation to prevent human rights violations and their fundamental role as guarantors of these rights. Thus, our recommendations involve concrete developments of the cross-cutting principles presented above to ensure the existence and proper operation of this fundamental framework. We present them below:

Social Participation: Meaningful social participation requires dedicated state officials, budget, processes, and planning, so that it is not merely an exercise in formal compliance without real substance.

Equality and non-discrimination: Complying with these principles also involves having the necessary structure for meaningful participation, especially to involve groups that have historically been discriminated against. It requires mobilizing diverse expertise within and outside state institutions to properly assess the system and the social context in which it will be implemented, and to address potential challenges.

problems. It requires continuous monitoring of implementation, with competent and responsible human oversight. Finally, it involves the production and analysis of human rights-oriented indicators on the system's implementation and its positive and negative impacts.

Due process/access to justice: This principle demands easily accessible, equitable, and effective judicial and administrative remedies, as well as adequate structures to investigate and sanction human rights violations resulting from state use of algorithmic systems, ensuring reparation and non-repetition. It is not uncommon for states' implementation of algorithmic systems to directly affect the recognition and exercise of essential rights without providing efficient and accessible appeal mechanisms.

Transparency/access to information: Requires routines and personnel to produce, organize, and disseminate information, both proactively and in response to requests, as well as an independent oversight body with sufficient and effective powers.

Privacy and data protection: These principles require that state institutions evaluating and implementing AI/ADM systems have adequate security infrastructure and expertise. The implementation of algorithm-based projects involving the processing of personal data should also be under the supervision of an independent data protection authority and state departments or officials within the program's responsible institutions who can serve as data protection officers (before whom individuals can exercise their data-related rights, such as access and rectification). Measures and procedures should also be in place to timely satisfy the rights of data subjects.

More on civic participation and transparency

Civic participation and public engagement are essential for designing and implementing algorithmic accountability policies. To ensure meaningful public participation, the following recommendations can be considered:

- I Conduct effective public interest consultations involving multiple stakeholders. To foster equity and inclusion, consider employing methods such as deliberative polling, citizen reference panels, citizen juries, and participatory budgeting.⁴³ A notable example is Oakland's Community Safety and Policing Ordinance, which requires extensive public hearings through established forums such as council meetings.

⁴³ <https://www.oecd.org/gov/open-government/eight-ways-to-institutionalise-deliberative-democracy.htm>

⁴⁴ <https://www.opengovpartnership.org/wp-content/uploads/2021/08/algorithmic-accountability-public-sector.pdf>

I Collaborate with marginalized communities and organizations representing marginalized groups to understand and eliminate the specific barriers they face. Equity and accessibility should be ensured by removing barriers to their participation, providing reasonable accommodations (e.g., sign language interpreters, indigenous language interpreters), compensating representatives, using accessible formats and communication channels, and funding and supporting capacity-building programs.

45

I Base AI-related decision-making processes on established ex-ante participatory rights frameworks, such as the UNECE Aarhus Convention.⁴⁶

I Demand transparency of AI systems and their procurement and deployment processes.⁴⁷ Many countries lack transparency in the procurement and deployment of AI technologies, with opaque contracting processes that fuel corruption and limit competition.⁴⁸ Public bodies must be aware of the type of information they share and how certain audiences access, trust, or use it. For example, the Amsterdam and Helsinki algorithm registries were specifically designed for critical audiences such as civil society, while France uses plain-language audiovisual explanations to inform affected communities and the public.⁴⁹

It is worth remembering that the inter-American participatory mechanisms in the context of indigenous and Afro-descendant communities provide valuable models and lessons for to promote meaningful participation in other contexts as well – especially Regarding what is meant by free, prior, and informed consultation with affected communities. Some of these are:

- Consultation is not a single act, but a process of dialogue in which clear, accessible and complete information is provided with sufficient time to allow for appropriate engagement⁵⁰;

⁴⁵https://gpai.ai/projects/responsible-ai/towardsrealdiversityandgenderequalityinai/towards-substantive-equality%20in-artificial-intelligence_Transformative-AI-policy-for-gender-equality-and-diversity.pdf

⁴⁶ <https://unece.org/environment-policy/public-participation/aarhus-convention/introduction> .

⁴⁷ https://ia.derechosdigitales.org/wp-content/uploads/2024/12/2024-LATAM-IA_en_el-Estado-ES.pdf

⁴⁸ https://ia.derechosdigitales.org/wp-content/uploads/2024/12/2024-LATAM-IA_en_el-Estado-ES.pdf ; https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/07/public-procurement-in-the-state-of-mexico_df343ad7/cc1da607-en.pdf

⁴⁹ <https://www.opengovpartnership.org/wp-content/uploads/2021/08/executive-summary-algorithmic-accountability.pdf>

⁵⁰ IACHR, Rights of Indigenous and Tribal Peoples to their Ancestral Lands and Natural Resources: Standards and Jurisprudence of the Inter-American Human Rights System, December 30, 2009, para. 285, and IACHR, Indigenous Peoples, Communities

- Consultation in good faith requires the absence of any type of coercion and must go beyond merely pro forma procedures⁵¹;
- Failure to take due account of the results of the consultation is contrary to the principle of good faith⁵²;
- The decisions resulting from the consultation process are subject to the higher administrative and judicial authorities, through appropriate and effective procedures, to assess their validity, relevance and the balance between the rights and interests at stake.⁵³

As mentioned, transparency is a key element in guaranteeing the right to access information, meaningful civic participation, and other rights directly implicated in the state's use of algorithmic systems. The EFF's report on the application of inter-American human rights standards in this context develops specific implications and obligations [regarding transparency](#), which can be consulted [in more detail](#). General recommendations include: _____

- Prioritize the development, acquisition, and deployment of open-source algorithm systems. Making the system's source code publicly available is a crucial transparency measure, which we urge States to adopt for systems that support consequential decisions about people's lives. In any case, a proper assessment of the system before and during its implementation requires that the State institution responsible for its implementation and independent experts have access to and analyze its source code;
- Make system specification documentation publicly available, establishing this as a prerequisite for developers and suppliers to contract with state institutions;
- Adopt only models that are interpretable⁵⁴ or that follow an explainability-by-design approach⁵⁵.

Afro-descendants and Natural Resources: Protection of Human Rights in the Context of Extractive, Exploitation, and Development Activities, December 31, 2015, para. 108.

⁵¹ IACHR, Rights of Indigenous and Tribal Peoples to Their Ancestral Lands and Natural Resources, para. 318.

⁵² IACHR, Rights of Indigenous and Tribal Peoples to Their Ancestral Lands and Natural Resources, para. 325.

⁵³ IACHR, Rights of Indigenous and Tribal Peoples to Their Ancestral Lands and Natural Resources, para. 328.

⁵⁴ Rudin, C. (2019). Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nat Mach Intell* 1, 206-215. See also Hamon, R., Junklewitz, H., & Sanchez, I. (2020), p. 13 (section 3.2.4 Interpretable models vs. post-hoc interpretability).

⁵⁵ Hamon, R., Junklewitz, H., & Sanchez, I. (2020), p. 24.

c) On the use of AI in law enforcement.

1. Digital training and skills:

- o Design continuous training programs in AI, data analysis and cybersecurity.
- o Create multidisciplinary teams (lawyers + technologists) to develop tools aligned with human rights.

2. Transparency and accountability:

- o Publish regulations, privacy notices and AI protocols on portals of transparency.
- o Establish clear mechanisms to audit algorithms and avoid bias.

3. Protection of human rights, including data protection:

- o Conduct human rights impact assessments before implement AI.
- o Guarantee informed consent and ARCO rights in the treatment of data.

4. Cybersecurity and document management:

- o Adopt [ISO 2700156 standards](https://www.iso.org/en/standard/27001) to protect sensitive information.
- o Digitize physical files and create secure backup protocols.

⁵⁶ <https://www.iso.org/en/standard/27001>